Biodiversity Data Journal 13: e136243
doi: 10.3897/BDJ.12.e136243

Data Paper

Mobilising marine biodiversity data: a new malacological dataset of Italian records (Mollusca)

Arianna Giannini‡, Massimo Appolloni§, Luigi Romani|,¶, Marco Oliverio‡

‡ Department of Biology and Biotechnologies "Charles Darwin", Sapienza University, Rome, Italy
§ Civic Museum of Zoology, Rome, Italy
| Institute of Systematics, Evolution, Biodiversity, Muséum national d'Histoire naturelle, Paris, France
¶ Via delle Ville, 79 - 55012 Capannori, Lucca, Italy

Corresponding author: Arianna Giannini (arianna.giannini@uniroma1.it)

Academic editor: Dimitris Poursanidis

Received: 03 Sep 2024 | Accepted: 09 Dec 2024 | Published: 28 Feb 2025

Citation: Giannini A, Appolloni M, Romani L, Oliverio M (2025) Mobilising marine biodiversity data: a new malacological dataset of Italian records (Mollusca). Biodiversity Data Journal 13: e136243. https://doi.org/10.3897/BDJ.12.e136243

Abstract

Background

The location and palaeoceanographic history of the Mediterranean Sea make it a biodiversity hotspot, prompting extensive studies in this region. However, despite the marine biodiversity of this area being apparently widely studied, a large amount of distributional data for Mediterranean taxa is still unpublished or scattered in various sources and formats, causing severe limitations to their potential reuse. This emerges as a particularly thorny issue for highly biodiverse and neglected taxa, such as invertebrates. The mobilisation of these frozen data through a process of standardisation and georeferencing could potentially support biodiversity research and conservation. The aim of this work is to provide a standardised pipeline to integrate these dispersed data, focusing on the Italian waters of the Mediterranean Sea and using molluscs as target taxa. Data were gathered from two main sources: published literature and Natural History Collections. The harmonisation process involved three key steps: 1) terminology and structure standardisation; 2) taxonomy updating and 3) georeferencing. Our efforts yielded over 44000 standardised records of mollusc species from Italian seawaters. These records encompassed primary biodiversity data from newly-digitised specimens

© Giannini A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

owned by 11 different institutions and private collectors, as well as secondary biodiversity data extracted from 311 published studies.

New information

This work is the first attempt to mobilise the available distributional information of Italian marine mollusc species from Natural History Collections and literature, converting the retrieved data into point-occurrence records through standard protocols, thus creating a FAIR (Findable, Accessible, Interoperable and Reusable) dataset collating these records from Italian marine sectors.

Keywords

marine, mollusca, big data, biodiversity conservation, Natural History Collections, occurrences, Italy

Introduction

Human impact on natural ecosystems is leading to several changes in global biodiversity structure and distribution. These changes include the loss of a portion of species large enough to suggest a sixth mass extinction, especially when coupled with the rate at which these extinctions are occurring (Cowie et al. 2022). This situation underscores the urgent need to enhance the quality and quantity of knowledge about species distribution over space and time, which is essential for making sustainable management decisions at various scales, from global to regional. In addition to primary biodiversity data gathered directly from fieldwork, another valuable source for enhancing our understanding of species distribution comes from already published data. However, this information is often published in formats that hinder easy reuse, such as species lists, non-standardised tables or occurrences reported only textually in the article body. The mobilisation of this information in formats that can be integrated and reused produces the so-called secondary biodiversity data (Marques et al. 2024). The possibility to integrate biodiversity data from various sources and formats (e.g. Natural History Collections (NHCs), literature, research and citizen-science programmes) can link the multiple domains of which biodiversity science is composed, such as taxonomy, ecology, conservation and socio-economic realms, filling gaps and offering new perspectives and opportunities (Heberling et al. 2021). With the 21th century, various organisations and facilities promoting the collection and integration of opportunistic biodiversity data have been created (e.g. Edwards et al. (2000)) and, to date, an increasing number of research and conservation programmes rely on point-occurrence records stored in this type of repositories. Although we are in the middle of the big data revolution (Runting et al. 2020), collation efforts remain uneven across geographical regions and taxa. Occurrences are biased towards regions and groups with more resources and a longer tradition of natural history studies (García-Roselló et al. 2023), as well as more accessible areas, resulting in marine

systems being less explored (Luypaert et al. 2019). Societal preferences play a crucial role in selecting the taxonomic groups on which to focus funding for data collection (Troudet et al. 2017) and public's *willingness to pay* is often centred on charismatic taxa, while invertebrates are some of the most neglected ones (Ressurreição et al. 2012), despite being amongst the most diverse and abundant groups. Those factors (with marine invertebrates data remaining frozen in their repositories) are currently impeding an unbiased prioritisation of both global and regional conservation strategies. In this context, in Italy, marine molluscs are often amongst the groups least considered in the development and management of conservation strategies, with the Habitats Directive (1992/43/EEC) deeming only 0.2% of Italian species worthy of conservation and a further 0.6% of species protected solely by the Bern and Barcelona Conventions. In addition to this, only 2% of the Mediterranean malacofauna is assessed according to Red List criteria and, of this percentage, 43% of the species are still Data Deficient (IUCN 2024). Fifteen of the 17 Mediterranean species of conservation concern according to international agrreements were recently assessed according to the IUCN Red List Assessment criterion A3C (population reduction projected to be met in the future based on a decline in area of occupancy) (Schultz et al. 2023), but there is still a long way to go to gain an overall knowledge of the conservation status of the Italian marine malacofauna. As early as 2000, Scotti and Chemello (2000) attempted to raise conservationists' interest, noting that, at that time, no Italian marine mollusc species was assessed in the IUCN Red List, partly attributing this to the little international recognition given to Mediterranean researchers in this period. Today, 24 years later, there is still a lack of information on the conservation status of most of the Italian malacofauna and, consequently, a lack of adequate conservation measures. In this context, the present work proposes to start filling this knowledge gap by building a FAIR dataset by integrating distributional information of marine mollusc species reported in Italy and making them usable as point-occurrence records, from newly-digitised specimens preserved in both public and private Natural History Collections, as well as mobilising secondary biodiversity data from published studies.

General description

Purpose: The present work aims at collecting and making usable in the form of point-occurrences the distributional data of marine mollusc species reported in Italy, by integrating via harmonisation and georeferencing processes both primary (i.e. newly-digitised specimen from public and private Natural History Collections) and secondary biodiversity data (i.e. non-databased spatial information of species reported in publicly-accessible papers). The dataset concept can be visualised in Fig. 1.

Sampling methods

Sampling description: Data were gathered from two main sources: literature and Natural History Collections (NHCs). To collect literature data, a comprehensive search was performed on the public databases Scopus and Web of Science. In addition to this, we

also searched data from journals specialised on Mediterranean marine fauna, namely *Iberus* and all the volumes of both journals of the Italian Society of Malacology (*Società Italiana di Malacologia*, SIM): *Bollettino Malacologico* and *Alleryana*. Since until the publication of the *Checklist of the Italian Fauna* (Bedulli et al. 1995a, Bedulli et al. 1995b, Bedulli et al. 1995c, Bello 1995, Bodon et al. 1995, Manganelli et al. 1995), no unified standard existed for Italian molluscan taxonomy and nomenclature - and verifying the accuracy of identifications reported in literature would have been difficult without direct check of the actual specimens - the literature search was restricted to publications issued after the first edition of the *Checklist of the Italian Fauna*. Species distribution information published in various formats (e.g. data tables in supplementary materials or within the paper, species lists, statements reporting the species occurrence) were considered as potential raw data. Papers with data already published in public databases were excluded. In order to avoid collecting the same record several times, only papers with new data were considered (i.e. new data derived from a dedicated sampling or non-new data published for the first time). In Table 1, search information are summarised. The full list of papers from which data were gathered is provided in Suppl. material 1. Data from NHCs were collected by direct request to private collectors (Giannini and Oliverio 2022) and institutions. The number of raw records initially received from each NHC owner and the final number of standardised records are listed in Suppl. material 2, together with the code used to identify the institution - derived from the Global Registry of Scientific Collections (Liu et al. 2024) - or private collector in the dataset. From both sources, records were included in the dataset if at least the occurrence locality and a taxonomic identification at the genus level (or more specific) were stated.

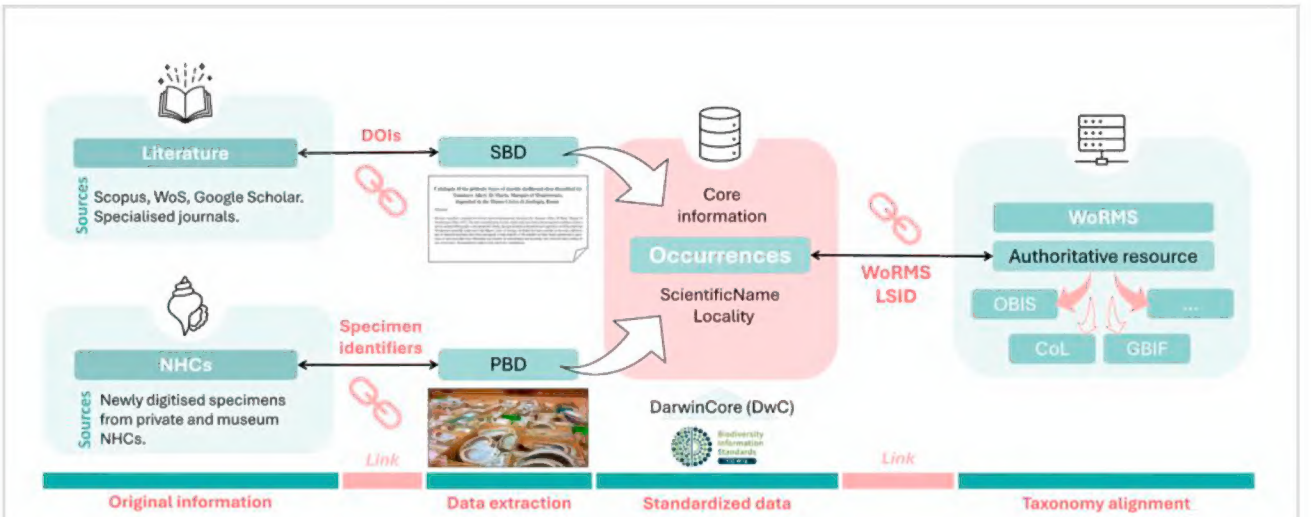

Figure 1. doi

Dataset concept. The dataset's core information consists of species occurrences. Occurrences are extracted from two sources: literature and NHCs. From the original source, records are converted into Primary or Secondary Biodiversity Data (respectively, PBD and SBD) through a process of standardisation and, when necessary, georeferencing. The scientific names of species are aligned with the World Register of Marine Species (WoRMS) nomenclature so that they are comparable with other taxonomic database resources (e.g. Ocean Biodiversity Information System: www.obis.org, Catalogue of Life: www.catalogueoflife.org, Global Biodiversity Information Facility: www.gbif.org).

Table 1.

Literature data search constraints, dates and number of results obtained for each consulted source.

Source	Search constraints	Start date	N° of results
Scopus	TITLE-ABS-KEY (marine AND mollusca AND italy)	16/05/2023	306
Web of Science	((ALL=(marine)) AND ALL=(mollusca)) AND ALL=(italy)	05/07/2023	442
Google Scholar	MARINE+MOLLUSCA+ITALY source:*Iberus*, from 1995 to 2023.	20/10/2023	33
SIM	All volumes of *Bollettino Malacologico* and *Alleryana* from 1995 to 2023.	24/11/2023	567
Total			**1348**

Quality control: To remove human-readable leading, trailing, double spaces and non-printable characters, the entire dataset was run through the Excel TRIM, CLEAN and SUBSTITUTE functions. Carriage returns were checked and removed using Notepad++ software (Ho 2011). To define the consistency and quality of the collected occurrences, a series of cleaning filters and manual checks were applied, which are described in detail in the next section (i.e. Step Description). To establish the quality of the taxonomic data, the resulting list of species in the dataset was extracted and compared with the *Checklist of Italian Fauna* (Renda et al. 2022). The species present in the dataset, but not included in the latter, were checked by expert taxonomists (under the supervision of Marco Oliverio and Luigi Romani), who solved problems and helped remove dubious records. The records underwent two geographic quality checks: the first one pre-georeferencing (Step description section, second step) - performed using the point-radius method (Wieczorek et al. 2010) - in which all records with too vague textual location were removed; and the second one post-georeferencing, in which all records with an uncertainty radius > 5000 m were removed (Step description section, seventh step). The double spatial filtering was necessary as high levels of uncertainty in georeferencing can create problems for spatial analyses (Reside et al. 2011). On the other hand, removing data with high spatial uncertainty can excessively reduce the sample size and, thus, the accuracy of the analyses (Smith et al. 2023) for which the data can be used. The uncertainty limit of 5000 m was chosen as a compromise between the completeness of the dataset and the quality and usability of the data. Some records from private NHCs refer to shells collected by fishermen and it was decided to keep them in the dataset, only if the fishing area/location was recorded and the geographical reliability of such records was acceptable, specifying that the collector is a fisherman in the *dwc:recordedBy* column to ensure transparency. In these cases, raw records are frequently accompanied by two locations: the port of arrival of the fishing vessel or the place where the collector acquired the specimen and the recorded collection area. If only the location of acquisition (and not the actual collection area) was explicitly indicated, the record was excluded from the dataset. This approach

was consistent with the treatment applied to specimens obtained through purchase or as gifts, if only the acquisition location was provided. Like all other records, locations that were too vague were excluded during the double geographical quality check.

Step description: 1. Firstly, data were merged and formatted in a Darwin Core scheme (Darwin Core Task Group 2009), using the Biodiversity Data Cleaning toolkit package (Ribeiro et al. 2022) in R (R Core Team 2022, ver. 4.2.2).

2. With the same package, a first filter was performed to clean the dataset from duplicates and records lacking essential information (i.e. identification or locality/coordinates). Then, data were manually filtered to retrieve records that were: out of scope (i.e. occurrences outside the Italian Marine Exclusive Economic Zone, fossils, non-marine species), too vague (i.e. broad locality, specimens with a higher level of identification than the genus) or dubious (dubious locality, ambiguous and/or unclear identification). All these cleaning steps have been consolidated into the "Invalid Records Filter" block in Fig. 2.

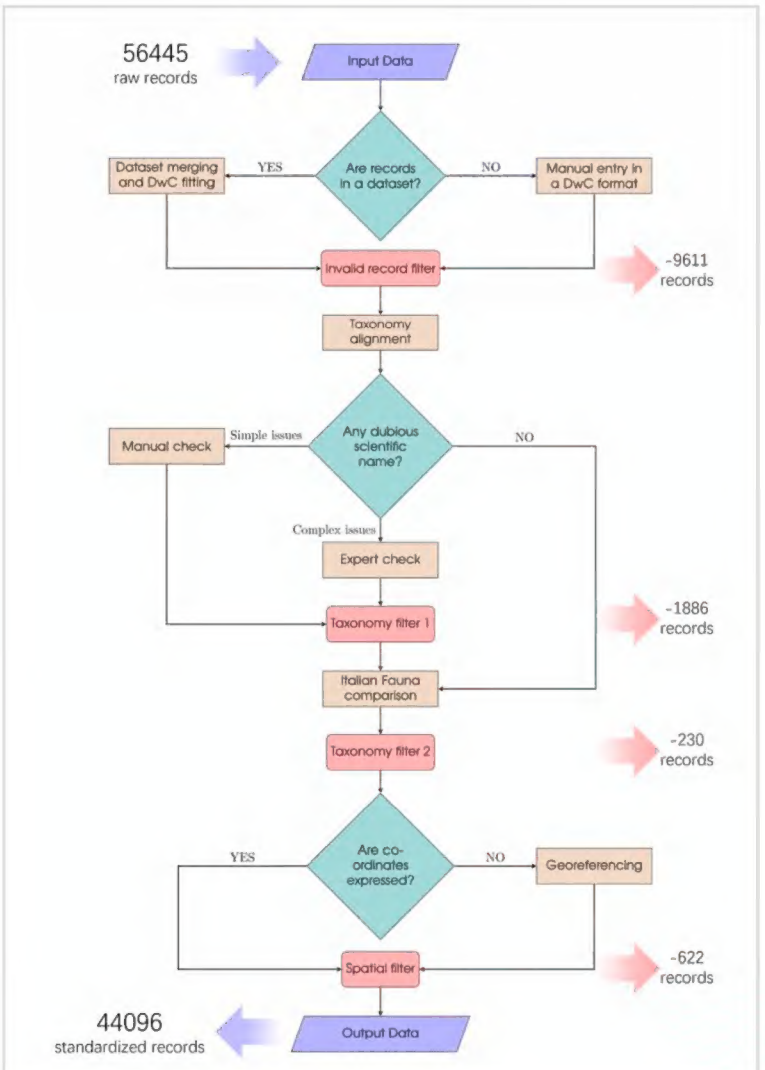

Figure 2. doi

Flow chart of the standardisation pipeline, with the number of input records and the number of discarded ones at each step. With the term "invalid", we identify all records out of scope or with problems (i.e. duplicates, records lacking essential information, fossils, non-marine species, specimens with a higher level of identification than the genus, ambiguous and/or unclear identification, records with dubious locality, occurrences outside study area or with broad locality).

3. Taxonomy was aligned to the one proposed by the World Register of Marine Species (WoRMS Editorial Board 2024) using the taxon-match Life Watch webservice (Fuentes and Fiore 2014), also extracting the WoRMS Life Science Identifiers (Vandepitte et al. 2015) for each valid scientific name to trace as far as possible rehashes of taxonomy, which in marine molluscs are quite common, especially through molecular evidence (e.g. Chiappa et al. (2022), Furfaro et al. (2022), Nocella et al. (2024)).

4. The remaining dubious taxonomy that was not automatically validated was checked manually and then submitted to experts, which resulted in the removal of other records with dubious identification.

5. Open Nomenclature (ON) qualifiers (Richter 1948, Matthews 1973, Bengtson 1988) were used to set uncertainty and provisional statuses for taxonomic identifications, applied following Sigovini et al. (2016) guidelines for ON terms harmonisation.

6. Subsequently, records were classified in seven different groups based on the type of the geographic information they had, in order to georeference them by the most appropriate method. Georeferencing was performed following the point-radius method (Wieczorek et al. 2010) and the Zermoglio et al. (2020) protocol, using GEOLocate web-based collaborative client (Rios and Bart 2010) and QGIS (QGIS.org 2024, ver. 3.22.13). Each final processed record has associated coordinates expressed in WGS84 decimal degrees and an uncertainty measure in metres. Table 2 details the method and, where applicable, protocol of georeferencing applied to each of the seven original geographic information classes. The GEBCO_2022 global terrain model was used to georeference depth data correctly (GEBCO Bathymetric Compilation Group 2022).

Table 2.

Description of the seven types of geographic information contained in the original records and georeferencing protocols and methods applied in each case. This information can be found in the dataset column *dwc:georeferenceRemarks*.

Type	Description	Georeference method and protocol	N° of records
Corrected	The original coordinates placed the raw record on land.	The record was moved to the position at sea nearest to the one defined by the original coordinates. A standard uncertainty of 100 m was assigned to the record.	748
Depth driven	The original geographical information was provided as a textual locality and the depth at which the specimen was found.	The record was georeferenced following the Zermoglio et al. (2020) protocol.	21144
Distance driven	The original geographical information was provided as the distance from a locality.	The record was georeferenced following the Zermoglio et al. (2020) protocol.	262

Type	Description	Georeference method and protocol	N° of records
Exact	The raw record already had exact coordinates provided in some format.	Where necessary, coordinates were converted to WGS84 decimal degrees. A standard uncertainty of 50 m was assigned to the record.	9012
Locality approximation	The original geographical information was provided as a textual locality, without depth or distance from the coast.	The record was georeferenced following the Zermoglio et al. (2020) protocol.	9430
Map approximation	The raw record was geographically positioned through a visual representation (e.g. map, satellite imagery) in the original source.	The record has been mapped trying to recreate as closely as possible the position represented in the original source. A standard uncertainty of 500 m was assigned to the record.	2894
Route	The raw record was geographically defined by the coordinates of the start and end of the route taken during the event in which the specimen was found.	The record was georeferenced following the Zermoglio et al. (2020) protocol.	606

7. During the georeferencing process it was possible to remove other data occurring outside study boundaries. We then excluded records with >5000 m of uncertainty radius.

8. As raw temporal data from NHCs arrived in various formats, this information was handled with the R package lubridate (Grolemund and Wickham 2011) and converted to ISO 8601 format. The temporal information collected spans various degrees of resolution, from the exact date to time ranges between years. We decided to mantain also records without temporal information. No hourly data were collected. Standardision and cleaning pipeline main steps are reported in Fig. 2.

Geographic coverage

Description: Collected data occurred within the Italian Exclusive Economic Zone (EEZ), that consists of a marine area of 538,216 km^2 (Khalfallah et al. 2020) situated in the Mediterranean Sea. This area is divided into nine biogeographical sectors based on the distribution of coastal taxa and the presence of physical, hydrological and physiological barriers (Bianchi 2004, Giacobbe and Oliverio 2024): Ligurian Sea, Northern Tyrrhenian, Southern Tyrrhenian and Messina Strait on the west coast of Italy, Northern Adriatic, Southern Adriatic and Mid-Adriatic on the east coast and Southern Mediterranean and Ionian Sea in the southern part. This subdivision has already been used to study the distribution of both marine plant and animal species in the marine flora and fauna checklist of the Società Italiana di Biologia Marina (Relini 2008, Relini 2010) and for the latest version of the Italian checklist of the marine malacofauna (Renda et al. 2022) and has been proposed as a standard in the Marine Regions Gazetteer (Flanders Marine Institute 2024). Sectors with the highest number of recorded occurrences are the

Tyrrhenian Sea (with 10480 occurrences for the Southern area and 10076 for the Northern one), followed by the Ionian Sea (7284), Ligurian Sea (5384), Northern Adriatic (2652), Mid-Adriatic (2647), Messina Strait (2212), Southern Mediterranean (1776) and Southern Adriatic (1585). Since sector areas are very diverse, from the smallest (the Messina Strait) to the largest (the Northern Tyrrhenian Sea) data coverage for each sector may vary depending on its size. Despite this general trend, the Strait of Messina, which covers only 0.2% of the Italian sea surface, contains about 5.0% of the dataset's occurrences, proving to be a particularly well-studied area (Fig. 3). The distribution of collected, cleaned and georeferenced occurrences is shown in Fig. 4.

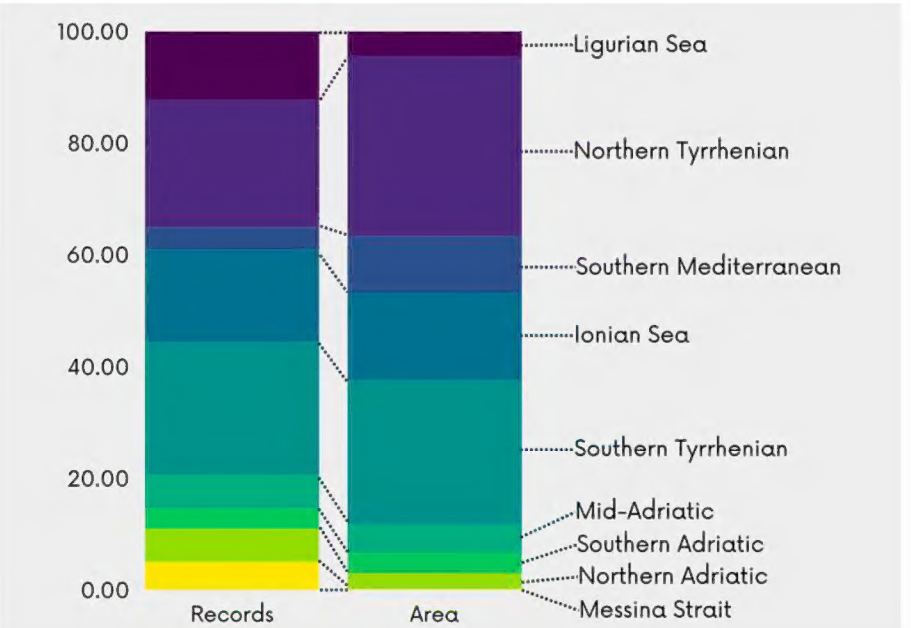

Figure 3. doi

Data abundance (%) per biogeographical sector compared with its extent (%) in the Italian Marine Exclusive Economic Zone.

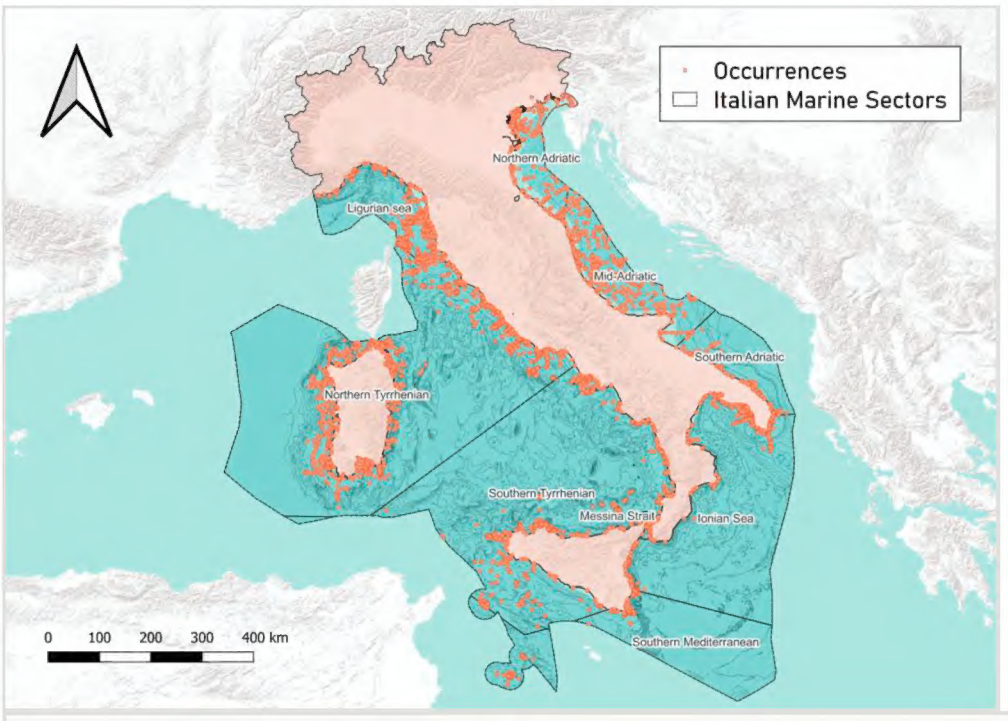

Figure 4. doi

Distribution of collected occurrences.

Coordinates: 35.06440614922465 and 45.80891370810167 Latitude; 5.889722222129848 and 18.99523827942329 Longitude.

Taxonomic coverage

Description: The dataset includes 44096 occurrences of 1513 Italian marine mollusc species, covering 85% of the Italian malacofauna and six out of the eight classes reported in Italy (Renda et al. 2022): Gastropoda (1110 species, 32832 occurrences), Bivalvia (326 species, 9410 occurrences), Polyplacophora (28 species, 820 occurrences), Cephalopoda (35 species, 682 occurrences), Scaphopoda (12 species, 331 occurrences) and Monoplacophora (2 species, 21 occurrences) (Table 3). Of the specimens reported in the dataset, 97% are identified at least at species level, while the remaining part is identified at genus level. The complete list of species can be viewed in Suppl. material 3, while Suppl. material 4 contains annotations on the inclusion of some ambiguous taxa (i.e. taxonomic entities that are difficult to interpret/not completely resolved, possible fossil specimens and occasional alien species for the Italian fauna) in the dataset.

Table 3.

Number of genera, species, subspecies and occurrences per class.

Class	Genus	Species	Subspecies	Occurrences
Bivalvia	207	326	1	9410
Cephalopoda	28	35	0	682
Gastropoda	457	1110	7	32832
Monoplacophora	2	2	0	21
Polyplacophora	12	28	0	820
Scaphopoda	7	12	0	331
Total	**713**	**1513**	**8**	**44096**

Taxa included:

Rank	Scientific Name
kingdom	Animalia
phylum	Mollusca
class	Bivalvia
class	Cephalopoda
class	Gastropoda
class	Monoplacophora

class	Polyplacophora
class	Scaphopoda
family	Acanthochitonidae
family	Acteonidae
family	Addisoniidae
family	Aegiridae
family	Aeolidiidae
family	Aglajidae
family	Akeridae
family	Alacuppidae
family	Amathinidae
family	Anabathridae
family	Anatomidae
family	Anomiidae
family	Aplysiidae
family	Aporrhaidae
family	Architectonicidae
family	Arcidae
family	Argonautidae
family	Arminidae
family	Assimineidae
family	Astartidae
family	Atlantidae
family	Barleeiidae
family	Basterotiidae
family	Bathysciadiidae
family	Borsoniidae
family	Brachioteuthidae
family	Bullidae
family	Bursidae
family	Cadlinidae

family	Caecidae
family	Calliostomatidae
family	Callistoplacidae
family	Callochitonidae
family	Calmidae
family	Calycidorididae
family	Calyptraeidae
family	Cancellariidae
family	Capulidae
family	Cardiidae
family	Carditidae
family	Carinariidae
family	Cassidae
family	Cavoliniidae
family	Cerithiidae
family	Cerithiopsidae
family	Chamidae
family	Charoniidae
family	Chauvetiidae
family	Chilodontaidae
family	Chitonidae
family	Chromodorididae
family	Cimidae
family	Cingulopsidae
family	Clathurellidae
family	Clavagellidae
family	Cliidae
family	Cocculinidae
family	Colloniidae
family	Colpodaspididae
family	Colubrariidae

family	Columbellidae
family	Conidae
family	Corbulidae
family	Cornirostridae
family	Coryphellidae
family	Costellariidae
family	Cranchiidae
family	Crassatellidae
family	Creseidae
family	Cuspidariidae
family	Cylichnidae
family	Cymatiidae
family	Cymbuliidae
family	Cypraeidae
family	Cystiscidae
family	Dendrodorididae
family	Dentaliidae
family	Diaphanidae
family	Discodorididae
family	Donacidae
family	Doridiidae
family	Dotidae
family	Dreissenidae
family	Drilliidae
family	Elachisinidae
family	Eledonidae
family	Ellobiidae
family	Embletoniidae
family	Entalinidae
family	Epitoniidae
family	Eratoidae

family	Eubranchidae
family	Eulimidae
family	Facelinidae
family	Fasciolariidae
family	Fionidae
family	Fissurellidae
family	Flabellinidae
family	Fusiturridae
family	Fustiariidae
family	Gadilidae
family	Gadilinidae
family	Galeommatidae
family	Gastrochaenidae
family	Glossidae
family	Glycymerididae
family	Goniodoridídae
family	Granulinidae
family	Gryphaeidae
family	Haliotidae
family	Halonymphidae
family	Haminoeidae
family	Hancockiidae
family	Hanleyidae
family	Heliconoididae
family	Hermaeidae
family	Heroidae
family	Hiatellidae
family	Histioteuthidae
family	Horaiclavidae
family	Hyalocylidae
family	Hyalogyrinidae

family	Hydrobiidae
family	Iravadiidae
family	Ischnochitonidae
family	Isognomonidae
family	Janolidae
family	Kelliellidae
family	Laonidae
family	Larocheidae
family	Lasaeidae
family	Lepetellidae
family	Lepetidae
family	Leptochitonidae
family	Limacinidae
family	Limapontiidae
family	Limidae
family	Limopsidae
family	Littorinidae
family	Loliginidae
family	Lottiidae
family	Lucinidae
family	Lyonsiellidae
family	Lyonsiidae
family	Mactridae
family	Malleidae
family	Malletiidae
family	Mangeliidae
family	Margaritidae
family	Marginellidae
family	Mathildidae
family	Mesodesmatidae
family	Mitridae

family	Mitromorphidae
family	Murchisonellidae
family	Muricidae
family	Myidae
family	Myrrhinidae
family	Mytilidae
family	Nassariidae
family	Naticidae
family	Neoleptonidae
family	Neopilinidae
family	Neritidae
family	Newtoniellidae
family	Noetiidae
family	Notodiaphanidae
family	Nuculanidae
family	Nuculidae
family	Octopodidae
family	Octopoteuthidae
family	Ocythoidae
family	Omalogyridae
family	Ommastrephidae
family	Onchidiidae
family	Onchidorididae
family	Onychoteuthidae
family	Orbitestellidae
family	Ostreidae
family	Otinidae
family	Ovulidae
family	Oxynoidae
family	Pandoridae
family	Parilimyidae

family	Patellidae
family	Pectinidae
family	Pediculariidae
family	Pendromidae
family	Peraclidae
family	Periplomatidae
family	Pharidae
family	Phasianellidae
family	Philinidae
family	Pholadidae
family	Phyllidiidae
family	Pinnidae
family	Pisaniidae
family	Piseinotecidae
family	Plakobranchidae
family	Planaxidae
family	Platyhedylidae
family	Pleurobranchaeidae
family	Pleurobranchidae
family	Polyceridae
family	Poromyidae
family	Potamididae
family	Pristiglomidae
family	Propeamussiidae
family	Psammobiidae
family	Pseudococculinidae
family	Pteriidae
family	Pterotracheidae
family	Pulsellidae
family	Pyramidellidae
family	Ranellidae

family	Raphitomidae
family	Retusidae
family	Rhizoridae
family	Ringiculidae
family	Rissoellidae
family	Rissoidae
family	Rissoinidae
family	Runcinidae
family	Samlidae
family	Scaliolidae
family	Scaphandridae
family	Scissurellidae
family	Scyllaeidae
family	Semelidae
family	Sepiidae
family	Sepiolidae
family	Siliquariidae
family	Siphonariidae
family	Skeneidae
family	Skeneopsidae
family	Solecurtidae
family	Solemyidae
family	Solenidae
family	Spondylidae
family	Tellinidae
family	Teredinidae
family	Tethydidae
family	Thraciidae
family	Thyasiridae
family	Thysanoteuthidae
family	Tjaernoeiidae

family	Tonicellidae
family	Tonnidae
family	Tornidae
family	Trapezidae
family	Tremoctopodidae
family	Trimusculidae
family	Trinchesiidae
family	Triphoridae
family	Tritoniidae
family	Triviidae
family	Trochaclididae
family	Trochidae
family	Truncatellidae
family	Tudiclidae
family	Turbinidae
family	Turritellidae
family	Tylodinidae
family	Umbraculidae
family	Ungulinidae
family	Vanikoridae
family	Velutinidae
family	Veneridae
family	Vermetidae
family	Verticordiidae
family	Vitrinellidae
family	Volvatellidae
family	Xenophoridae
family	Xylodisculidae
family	Xylophagaidae
family	Yoldiidae

Temporal coverage

Data range: 1814-1-01 - 2023-12-31.

Notes: The dataset includes both recent and historical data. Regarding NHCs data, pre-1950 records come from the historical collections held in the Civic Museum of Zoology of Rome (i.e. Monterosato, Meli and Piersanti NHCs), while post-1950 records are from private collections. For literature data, on the other hand, pre-1950 records come mainly from published catalogues of historical collections, while more recent records come from faunistic and ecological studies. The list of all literature and NHCs data sources can be found respectively in Suppl. material 1 and in Suppl. material 2. The number of records per year obtained from literature and from NHCs can be visualised in Fig. 5.

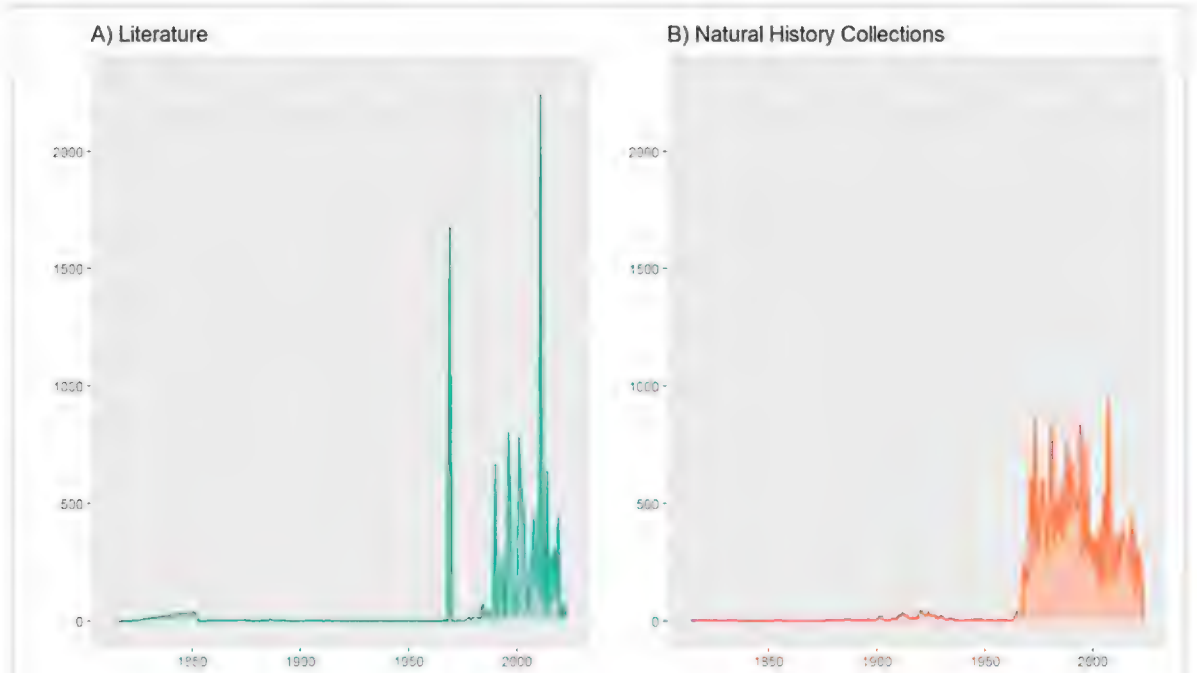

Figure 5. doi

Temporal coverage of the data, divided by data obtained through literature search (A) and from Natural History Collections (B).

Usage licence

Usage licence: Other

IP rights notes: **Creative Commons Attribution Non-Commercial (CC-BY-NC 4.0)**

Data resources

Data package title: Mobilising marine biodiversity data: a new malacological dataset of Italian records (Mollusca)

Resource link: https://www.gbif.org/dataset/e0370da7-b32b-414c-8a61-1d86125075f3

Alternative identifiers: https://cloud.gbif.org/eca/resource?r=mzur_sap_zoo_01

Number of data sets: 1

Data set name: Mobilising marine biodiversity data: a new malacological dataset of Italian records (Mollusca)

Data format: Darwin Core Archive

Column label	Column description
institutionID	An identifier for the institution having custody of the object(s) or information referred to in the record. For data collected from public institutions, the identifiers of the Global Registry of Scientific Collections (GRSciColl) were used.
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record. For data collected from public institutions, the identifiers of the Global Registry of Scientific Collections (GRSciColl) were used. For data collected from private collections, the following non-standard identifiers were used, marked with the prefix "PriColl_": Nofroni I. (PriColl_NI), Renda W. (PriColl_RW), Romani L. (PriColl_RL), Roncone F. (PriColl_RF), Russo P. (PriColl_RP), Tringali L. (PriColl_TL), Trono D. (PriColl_TD).
collectionCode	The name, acronym, coden or initialism identifying the collection or dataset from which the record was derived.
basisOfRecord	The specific nature of the data record.
catalogNumber	An identifier for the record within the dataset or collection.
recordedBy	A list (concatenated and separated) of names of people, groups or organisations responsible for recording the original dwc:Occurrence.
individualCount	The number of individuals present at the time of the dwc:Occurrence.
establishmentMeans	Statement about whether a dwc:Organism has been introduced to a given place and time through the direct or indirect activity of modern humans.
associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the dwc:Occurrence. This column was used to indicate the bibliographic source from which literature data was collected.
eventDate	The date or interval during which a dwc:Event occurred. Dates are expressed following the ISO 8601 standard.
year	The four-digit year in which the dwc:Event occurred, according to the Common Era Calendar.
higherGeographyID	The Getty Thesaurus of Geographic Names persistent identifier for the geographic region within which the dcterms:Location occurred.

higherGeography	The less specific geographic name of the information captured in the dwc:locality term.
waterBody	The name of the water body in which the dcterms:Location occurs. Names of the Italian marine biogeographical areas were used (Bianchi et al. 2004).
country	The name of the country in which the dcterms:Location occurs.
stateProvince	The name of the next smaller administrative region than country (i.e. region) in which the dcterms:Location occurs.
locality	The specific description of the place.
minimumDepthInMetres	The lesser depth of a range of depth below the local surface, in metres.
maximumDepthInMetres	The greater depth of a range of depth below the local surface, in metres.
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in dwc:geodeticDatum) of the geographic centre of a dcterms:Location.
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in dwc:geodeticDatum) of the geographic centre of a dcterms:Location.
geodeticDatum	The ellipsoid, geodetic datum or spatial reference system (SRS), upon which the geographic coordinates given in dwc:decimalLatitude and dwc:decimalLongitude are based.
coordinateUncertaintyInMetres	The horizontal distance (in metres) from the given dwc:decimalLatitude and dwc:decimalLongitude describing the smallest circle containing the whole of the dcterms:Location.
georeferencedBy	A list (concatenated and separated) of names of people, groups or organisations who determined the georeference (spatial representation) for the dcterms:Location.
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates and uncertainties.
georeferenceRemarks	Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalised in the method referred to in dwc:georeferenceProtocol.
identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the dwc:Identification. The standard terminology proposed by Sigovini et al. (2016) was followed.
identifiedBy	A list (concatenated and separated) of names of people, groups or organisations who assigned the dwc:Taxon to the subject.
scientificNameID	An identifier for the nomenclatural (not taxonomic) details of a scientific name. WoRMS LSID persistent identifiers were used.
scientificName	The full scientific name.
class	The full scientific name of the class in which the dwc:Taxon is classified.

order	The full scientific name of the order in which the dwc:Taxon is classified.
family	The full scientific name of the family in which the dwc:Taxon is classified.
genus	The full scientific name of the genus in which the dwc:Taxon is classified.
specificEpithet	The name of the first or species epithet of the dwc:scientificName.
infraspecificEpithet	The name of the lowest or terminal infraspecific epithet of the dwc:scientificName.
taxonRank	The taxonomic rank of the most specific name in the dwc:scientificName.
scientificNameAuthorship	The authorship information for the dwc:scientificName formatted according to the conventions of the applicable dwc:nomenclaturalCode.

References

- Bedulli D, Dell'Angelo B, Salvini-Plawen L (1995a) Caudofoveata, Solenogastres, Polyplacophora, Monoplacophora. In: Minelli A, Ruffo S, La Posta S (Eds) Check-list delle specie della fauna italiana. 13. Calderini, Bologna, 5 pp. [In Italian].
- Bedulli D, Castagnolo L, Ghisotti F, Spada G (1995b) Bivalvia, Scaphopoda. In: Minelli A, Ruffo S, La Posta S (Eds) Check-list delle specie della fauna italiana. 17. Calderini, Bologna, 21 pp. [In Italian].
- Bedulli D, Cattaneo-Vietti R, Chemello R, Ghisotti F, Giovine F (1995c) Gastropoda Opisthobranchia, Divasibranchia, Gymnomorpha. In: Minelli A, Ruffo S, La Posta S (Eds) Check-list delle specie della fauna italiana. 15. Calderini, Bologna, 24 pp. [In Italian].
- Bello G (1995) Cephalopoda. In: Minelli A, Ruffo S, La Posta S (Eds) Checklist delle specie della fauna italiana. 18. Calderini, Bologna, 5 pp. [In Italian].
- Bengtson P (1988) Open Nomenclature. Palaeontology 31 (1): 223-227.
- Bianchi CN (2004) Proposta di suddivisione dei mari italiani in settori biogeografici. Notiziario della Società Italiana di Biologia Marina 46: 57-59. [In Italian].
- Bodon M, L F, Giannuzzi-Savelli R, Giovine F, Giusti F, Manganelli G, Melone G, Oliverio M, Sabelli B, Spada G (1995) Gastropoda Prosobranchia, Heterobranchia Heterostropha. In: Minelli A, Ruffo S, La Posta S (Eds) Checklist delle specie della fauna italiana. 14. Calderini, Bologna, 60 pp. [In Italian].
- Chiappa G, Fassio G, Corso A, Crocetta F, Modica MV, Oliverio M (2022) How many abalone species live in the Mediterranean sea? Diversity 14 (12): 1107. https://doi.org/10.3390/d14121107
- Cowie R, Bouchet P, Fontaine B (2022) The sixth mass extinction: fact, fiction or speculation? Biological Reviews 97 (2): 640-663. https://doi.org/10.1111/brv.12816
- Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG). URL: http://www.tdwg.org/standards/450
- Edwards J, Lane M, Nielsen E (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. Science 289 (5488): 2312-2314. https://doi.org/10.1126/science.289.5488.2312
- Flanders Marine Institute (2024) Marine Regions. www.marineregions.org

- Fuentes D, Fiore N (2014) The LifeWatch approach to the exploration of distributed species information. ZooKeys 463: 133-148. https://doi.org/10.3897/zookeys.463.8397
- Furfaro G, Schreier C, Trainito E, Pontes M, Madrenas E, Girard P, Mariottini P (2022) The sea slug *Doriopsilla areolata* Bergh, 1880 (Mollusca, Gastropoda) in the Mediterranean sea: another case of cryptic diversity. Diversity 14 (4): 297. https://doi.org/10.3390/d14040297
- García-Roselló E, González-Dacosta J, Lobo J (2023) The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. Biological Conservation 283: 110118. https://doi.org/10.1016/j.biocon.2023.110118
- GEBCO Bathymetric Compilation Group (2022) GEBCO_2022 Grid. NERC EDS British Oceanographic Data Centre NOC.
- Giacobbe S, Oliverio M (2024) The conventional limits of the marine biogeographical Sector 4, for the new Checklist of the Italian Fauna. Biogeographia – The Journal of Integrative Biogeography 39 (2). https://doi.org/10.21426/b6.39863
- Giannini A, Oliverio M (2022) From shells in house cabinets to structured data for research: the mobilization of frozen biodiversity data in Italy. Biodiversity Information Science and Standards 6: e93584. https://doi.org/10.3897/biss.6.93584
- Grolemund G, Wickham H (2011) Dates and times made easy with lubridate. Journal of Statistical Software 40 (3): 1-25. https://doi.org/10.18637/jss.v040.i03
- Heberling JM, Miller J, Noesgaard D, Weingart S, Schigel D (2021) Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences 118 (6): e2018093118. https://doi.org/10.1073/pnas.2018093118
- Ho D (2011) Notepad++. 8.7.2. URL: https://notepad-plus-plus.org/
- IUCN (2024) The IUCN Red List of Threatened Species, Version 2024-1. https://www.iucnredlist.org. Accessed on: 2024-7-18.
- Khalfallah MC, Piroddi C, Brown SL, Noël T, Cashion B, Derrick D, Dunstan E, Page G, Tsui A, Pierucci B, et al. (2020) Mediterranean: updated catch reconstructions to 2018. In: Derrick B, Khalfallah M, Relano V, Zeller D, Pauly D (Eds) Updating to 2018 the 1950-2010 marine catch reconstructions of the Sea Around Us: Part I – Africa, Antarctica, Europe and the North Atlantic. 28 (5). Fisheries Centre Research Report.
- Liu MJ, Shrestha L, Grosjean M, Høfft M, Gonzalez M, Robertson T (2024) An Introduction to the Global Registry of Scientific Collections (GRSciColl). Biodiversity Information Science and Standards 8 https://doi.org/10.3897/biss.8.134257
- Luypaert T, Hagan J, McCarthy M, Poti M (2019) Status of marine biodiversity in the Anthropocene. YOUMARES 9 - The Oceans: Our Research, Our Future57-82. https://doi.org/10.1007/978-3-030-20389-4_4
- Manganelli G, Bodon M, Favilli L, Giusti F (1995) Gastropoda Pulmonata. In: Minelli A, Ruffo S, La Posta S (Eds) Checklist delle specie della fauna italiana. 16. Calderini, Bologna, 60 pp. [In Italian].
- Marques N, Soares CDdM, Casali DdM, Guimarães E, Fava F, Abreu JMdS, Moras L, Silva LGd, Matias R, Assis RLd, Fraga R, Almeida S, Lopes V, Oliveira V, Missagia R, Carvalho E, Carneiro N, Alves R, Souza-Filho P, Oliveira G, Miranda M, Tavares VdC (2024) Retrieving biodiversity data from multiple sources: making secondary data standardised and accessible. Biodiversity Data Journal 12 https://doi.org/10.3897/bdj.12.e133775

- Matthews S (1973) Notes on Open Nomenclature and on synonymy lists. Palaeontology 16: 713-719.
- Nocella E, Fassio G, Zuccon D, Puillandre N, Modica MV, Oliverio M (2024) From coral reefs into the abyss: the evolution of corallivory in the Coralliophilinae (Neogastropoda, Muricidae). Coral Reefs1-18. https://doi.org/10.1007/s00338-024-02537-1
- QGIS.org (2024) QGIS Geographic Information System. 3.22.13. QGIS Association. URL: https://www.qgis.org/
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing. 4.2.2. Vienna, Austria. URL: https://www.R-project.org/
- Relini G (2008) Checklist della Flora e della Fauna dei mari italiani (I parte). 15. Biologia Marina Mediterranea, 457 pp. [In Italian].
- Relini G (2010) Checklist della Flora e della Fauna dei mari italiani (II parte). 17. Biologia Marina Mediterranea, 517 pp. [In Italian].
- Renda W, Amati B, Bogi C, Bonomolo G, Capua D, Dell'Angelo B, Furfaro G, Giannuzzi Savelli R, La Perna R, Nofroni I, Pusateri F, Romani L, Russo P, Smriglio C, Tringali LP, Oliverio M (2022) The new Checklist of the Italian Fauna: marine Mollusca. Biogeographia – The Journal of Integrative Biogeography 37 (1): ucl005. https://doi.org/10.21426/b637156028
- Reside A, Watson I, VanDerWal J, Kutt A (2011) Incorporating low-resolution historic species location data decreases performance of distribution models. Ecological Modelling 222 (18): 3444-3448. https://doi.org/10.1016/j.ecolmodel.2011.06.015
- Ressurreição A, Gibbons J, Kaiser M, Dentinho TP, Zarzycki T, Bentley C, Austen M, Burdon D, Atkins J, Santos R, Edwards-Jones G (2012) Different cultures, different values: the role of cultural variation in public's WTP for marine species conservation. Biological Conservation 145 (1): 148-159. https://doi.org/10.1016/j.biocon.2011.10.026
- Ribeiro B, Velazco SJE, Guidoni-Martins K, Tessarolo G, Jardim L, Bachman S, Loyola R (2022) bdc: A toolkit for standardizing, integrating and cleaning biodiversity data. Methods in Ecology and Evolution 13 (7): 1421-1428. https://doi.org/10.1111/2041-210x.13868
- Richter R (1948) Einführung in die Zoologische Nomenklatur Durch Erläuterung der Internationalen Regeln. [Introduction to zoological nomenclature by explaining the international rules]. 2. Waldemar Kramer, Frankfurt, Germany. [In German].
- Rios N, Bart H (2010) GEOLocate. 3.22. Belle Chasse, LA: Tulane University Museum of Natural History.
- Runting R, Phinn S, Xie Z, Venter O, Watson JM (2020) Opportunities for big data in conservation and sustainability. Nature Communications 11 (1): 2003. https://doi.org/10.1038/s41467-020-15870-0
- Schultz L, Wessely J, Dullinger S, Albano PG (2023) The climate crisis affects Mediterranean marine molluscs of conservation concern. Diversity and Distributions 30 (3): e13805. https://doi.org/10.1111/ddi.13805
- Scotti G, Chemello R (2000) I Molluschi marini mediterranei degni di protezione: stato delle conoscenze e forme di tutela. Bollettino Malacologico 36 (1-4): 61-70. [In Italian].
- Sigovini M, Keppel E, Tagliapietra D (2016) Open Nomenclature in the biodiversity era. Methods in Ecology and Evolution 7 (10): 1217-1225. https://doi.org/10.1111/2041-210x.12594

- Smith A, Murphy S, Henderson D, Erickson K (2023) Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. Global Ecology and Biogeography 32 (3): 342-355. https://doi.org/10.1111/geb.13628
- Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Scientific Reports 7 (1): 9132. https://doi.org/10.1038/s41598-017-09084-6
- Vandepitte L, Vanhoorne B, Decock W, Dekeyzer S, Trias Verbeeck A, Bovit L, Hernandez F, Mees J (2015) How Aphia—the platform behind several online and taxonomically oriented databases—can serve both the taxonomic community and the field of biodiversity informatics. Journal of Marine Science and Engineering 3 (4): 1448-1473. https://doi.org/10.3390/jmse3041448
- Wieczorek J, Guo Q, Hijmans R (2010) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information Science 18 (8): 745-767. https://doi.org/10.1080/13658810412331280211
- WoRMS Editorial Board (2024) World Register of Marine Species. https://www.marinespecies.org
- Zermoglio P, Chapman A, Wieczorek J, Luna M, Bloom D (2020) Georeferencing Quick Reference Guide. Copenhagen: GBIF Secretariat. https://doi.org/10.35035/e09p-h128

Supplementary materials

Suppl. material 1: Sources from which literature data were obtained doi

Authors: Arianna Giannini
Data type: Table
Brief description: References of the 311 publications from which data were extracted and number of records obtained from each source.
Download file (46.21 kb)

Suppl. material 2: Sources from which NHCs data were obtained doi

Authors: Arianna Giannini
Data type: Table
Brief description: List of private collectors and institutions that own the 11 NHCs from which the data were collected, with initial number of raw records and final number of standardised records.
Download file (11.35 kb)

Suppl. material 3: Species list doi

Authors: Arianna Giannini, Luigi Romani, Marco Oliverio
Data type: Species list
Download file (549.86 kb)

Suppl. material 4: Notes on some taxa included in the dataset doi

Authors: Arianna Giannini
Data type: List
Brief description: This list contains annotations on some ambiguous taxa (i.e. taxonomic entities that are difficult to interpret/not fully resolved, possible fossil taxa and occasional alien species for the Italian fauna) that have nevertheless been included in the dataset.
Download file (11.79 kb)